DESIGN OF AN AI-BASED SMART WEARABLE DEVICE FOR VISUALLY IMPAIRED PEOPLE

Duc Long Vu, Duc-Hieu Nguyen, Duc Nguyen Phuong Phi and Hai-Chau Le

Posts and Telecommunications Institute of Technology, Hanoi, Vietnam

Abstract: The human visual system plays an important role in perceiving information related to the surrounding environment. Because visual cues provide more data than auditory information, visual cues are more effective than auditory cues when it comes to perceiving the information. However, in the case of visually impaired people, the lack of visual information will limit them in identifying information and people around them. Therefore, along with medical treatment technology, technology supports visually impaired people when moving is being studied and becoming urgent. Realizing the high applicability of recognition technology with artificial intelligence in helping the visually impaired, in this article, we propose and develop a smart wearable device that recognizes faces for visually impaired people applying deep learning techniques. The device was deployed on a compact, flexible hardware platform with low cost to identify and announce the name of the person identified to the visually impaired person through the voice. The results of experiment and evaluation show that the proposed device is extremely versatile and has achieved remarkable accuracy of 99.3% in a practical environment.

Keywords: IoT, Deep learning, Face detection, Face recognition, Convolution Neural Network.

I. INTRODUCTION

Nowadays, according to the World Health Organization, there are currently about 285 million people suffering from visual impairment worldwide [1]. Recent advances in modern technologies including Internet of things and artificial intelligence gradually enhance the intelligence of electronic devices to help simplify and make human life easier in a wide range of applications such as health care, education, finance, security and especially visually impaired person assistance [2, 3]. Many visually impaired people in the world are utilizing state-of-the-art technology to perform tasks in their daily lives. Such assistive technologies are commonly based on electronic devices equipped with sensors and processors capable of making “intelligent” decisions [4, 5]. Unfortunately, those critical assistive technologies are still out of reach for most of the visually impaired persons in developing countries like Vietnam where the number of blind and low-vision people (collectively referred to as the visually impaired persons) is more than two million while an estimated one-third of those have low incomes. Hence, development of low-cost and effective smart devices assisting visually impaired people is necessary.

In order to support visually impaired individuals, research and application of face detection and recognition have been attracting worldwide attention. At present, the development of intelligent algorithms, as well as image recognition and processing methods, have been improved to an almost absolute level of accuracy [6-8]. Various methods have been implemented for selection of the optimal application related to face recognition and face detection. In general, they are separated into two main classes that are machine learning [8-10] and deep learning [11-15]. Deep learning has emerged as an alternative method for machine learning, due to no requirement of human knowledge-based feature extraction and feature selection while offering an automatically learning feature, which is absent in machine learning [11, 12]. The face recognition and face detection performance of several deep learning models has been shown in many scientific reports with significant results [16-17]. In [16], researchers from Google proposed a deep learning model called FaceNet, which is known as the state-of-the-art deep learning model for face recognition with significant accuracy. This model uses a deep neural network extracting face features and using triplet-loss technique for training. Besides, FaceBook also proposes a deep learning facial recognition system called DeepFace [17], it consists of a nine-layer deep neural network which involves more than 120 million parameters and locally connected layers without weight sharing. A simple method using face thermal images and convolutional neural network (CNN) for face recognition is proposed in [18-20]. Although many face recognition applications are available for smartphone, tablet, security system, car, … the number of face recognition enable assistive systems, especially visually impaired person supporting devices, are limited [4, 5], and the price of almost such smart devices is still relatively high.

In this paper, we develop a low cost and simple wearable smart device system that enables face recognition by using an effective intelligent algorithm for assisting visually impaired people. The system is built on a multi-purpose computer platform, i.e. Raspberry Pi 4B, with necessary hardware modules including camera and voice interaction, and a high accurate face recognition software module. The smart wearable device is able to recognize familiar faces and instantly and accurately

Contact author: Hai-Chau Le
Email: Chaulh@ptit.edu.vn
Manuscript received: 19/6/2022, revised: 18/7/2022, accepted: 25/8/2022

inform their name. The system detects the faces from the images captured by the camera, then crops and converts the original face images into the appropriate grayscale images to speed up the process of the deep learning model implemented in the system. We extract features from facial images into a vector embedding. This vector embedding is used to train the classifier with a machine learning algorithm and put on the smart wearable device. Experiments have been done under various practical conditions. The obtained results prove that our developed system works properly and it can provide high accuracy, saying up to 99.3% in good light condition and at least 98.6% even in weak light condition.

II. AI-BASED SMART WEARABLE DEVICE FOR VISUALLY IMPAIRED PEOPLE

System Design

In this work, we target a small-scale, cost-effective, and highly accurate face recognizable smart wearable device using an intelligent algorithm for visually impaired people. In order to realize a low-cost and flexible device, we exploit multi-purpose hardware platforms, open-source software and mature AI technologies. Figure 1 shows the basic block diagram of our proposed smart wearable device. Our system is built with an open-source firmware, i.e. Raspbian [21], and a commercial off-the-shelf hardware, Raspberry pi 4 [22, 23, 24], and develop an intelligent algorithm to detect and recognize the familiar faces of the visually impaired person.

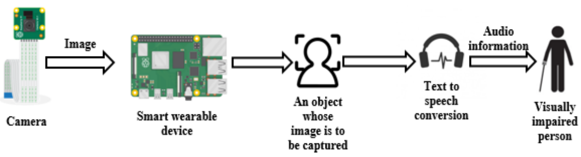


Figure 1. Block diagram of the proposed system

In figure 2, we illustrate the face recognition processing flow of our proposed smart wearable device.

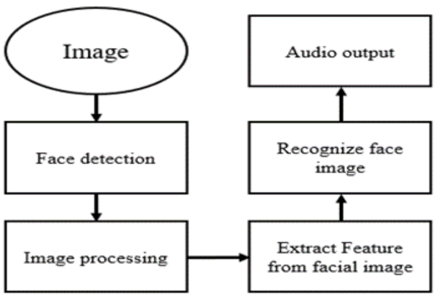


Figure 2. System flow of proposed smart device

Hardware Construction

Our developed device contains a Raspberry Pi Camera with 5 megapixels resolution and adjustable focus distance, a Raspberry Pi mini-computer, a headphone, and a power source. Raspberry Pi Camera captures a video frame and sends it to the main processor Raspberry Pi 4 where face recognition is performed. Raspberry Pi also produces associated audio names of the people who are recognized and sends it to the blind via headphones.

In our proposed smart device, we used a Raspberry Pi 4 Model B [22, 23, 24] as the main processor. This mini computing platform has 2 Gb Ram, 1.5 GHz CPU, a WiFi port 2.4 GHz and 5.0 GHz in IEEE 802.11ac standard, one port Gbps Ethernet and four USB ports [23]. The Raspberry pi 4 was powered by a backup charger which can supply 5 voltage direct current in about 12 hours.

Software Solution

1) Data

In this article we collected portrait images from 5 people in our laboratory for training our model. The dataset contains variations in facial poses such as front, left, right in different conditions of light. The accessories as glasses, hats also are included in this dataset. Total 1500 images had been used for training, validation and testing in our system. We collected 300 images for each person, which were then divided into 200 images for the training set, 70 images of the validation set, and 30 images of the test set. For image on training set, we were pre-processing image by detect human face in image using MTCNN model [25], then crop and convert facial image into grayscale, after that feed into deep learning model for face recognition.

2) Face detection method

In this paper, we used Multi-task Cascaded Convolution Neural Networks (MTCNN) for face detection task. MTCNN is a neural network which is known with high accuracy in face detection and alignment on images. It consists of 3 neural networks connected in a cascade [25,26]. Figure 3,4 and 5 show architecture of three convolutional neural networks in the MTCNN model.

In the first stage, a fully convolutional neural network is constructed called Proposal Network (P-Net). All candidate windows and their bounding box regression vectors are acquired. Then the bounding box vectors are used to calibrate the candidates. After That, a non-maximum suppression (NMS) is used for merging highly overlapping candidates. [26]

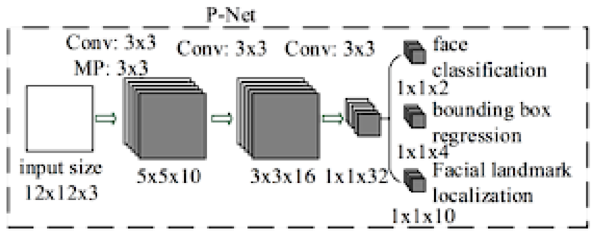


Figure.3 Structure of P-net

Then, all candidates obtained from P-net are sent to another CNN model, called Refine Network (R-net). In this stage, R-net removes a large number of non-face windows, performs calibrations with bounding boxes and merge candidates via NMS.

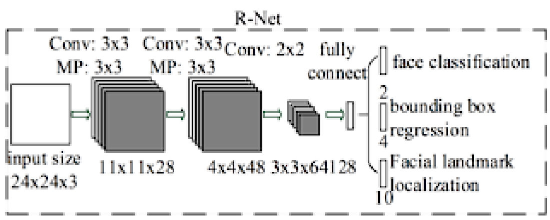


Figure 4. Structure of R-net

Finally, in the last stage the results from R-net are fed to Output Network (O-Net) which is an efficient CNN. This stage aims to optimize the result from R-net and out put five facial landmarks position and final bounding box.

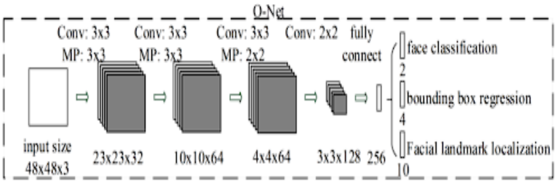


Figure 5. Structure of O-net

MTCNN have three tasks of training: face classification, bounding box regression and facial landmark localization. Each task has its own loss function. In face classification task or you may know it as a two-class classification problem, the cross-entropy loss is used for cost function.

$$L_i^{det} = -\left(y_i^{det}\, log\, log\,(p_i)\ + (1 - y_i^{det})(1 - p_i)\right) \quad (1)$$

Where p_i is the probability that indicates sample x_i being a face. The y_i^{det} is the ground truth label with two values {0,1}.

In bounding box regression, the loss function uses Euclidean loss for each sample x_i. For each sample we predict the offset between the candidate window and the nearest ground truth.

$$L_i^{box} = \|\hat{y}_i^{box} - y_i^{box}\|_2^2 \ (2) \qquad (2)$$

Where $\hat{y}_i^{box}$ represents the regression target created by P-net and y_i^{box} is the ground-truth coordinate. Euclidean loss is also used in Facial landmark localization task which is formulated as a regression problem just like previous task and we minimize the loss function as:

$$L_i^{landmark} = \|\hat{y}_i^{landmark} - y_i^{landmark}\|_2^2 \qquad (3)$$

Finally, the overall loss function is shown in formula (4).

$$f_{loss} = min \sum_{i=1}^{N} j \in \{det, box, landmark\} \alpha_j \beta_i^j L_i^j \quad (4)$$

Where N is the number of training examples and α represent the task importance. ($\alpha_{det} = 1, \alpha_{box} = 0.5, \alpha_{landmark} = 0.5$) is used in R-net and P-net. ($\alpha_{det} = 1, \alpha_{box} = 0.5, \alpha_{landmark} = 1$) is used in the output network O-net. The stochastic gradient descent is a useful method to train these CNNs.

3) Data pre-processing

After detecting the face in image using MTCNN, we crop the face image of the candidate and transfer it to a grayscale image for reduced dimension of input data then forward it to extract feature block which is shown in figure 6. This step is very important because reducing the dimension of the image which is represented in 3-D to 2-D helps us reduce computational cost in the face recognition task. Then, we resize gray scale images to 160x160 resolution to fit with input of FaceNet model in extract feature task.

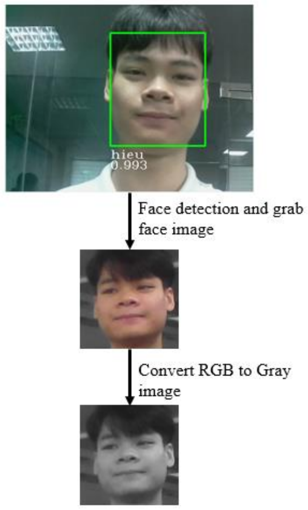


Figure 6. Face extraction

4) Feature extraction

In our smart device, we use the FaceNet model to extract some efficient features in the facial image of a candidate who is detected by the device. FaceNet is a deep neural network used for face recognition and it can be used as an embedding model for extracting features from an image of a person's face. The core of FaceNet is based on 2 types of CNNs including Zeiler & Fergus architecture [27] and Inception model [28] architecture based on GoogLeNet. Table I, II show structure of two CNN models above. The two architectures different in the number of parameters used and Float Point Operations Per Second (FLOP).

TABLE I. THE ZEILER & FERGUS NETWORK ARCHITECTURE

layer	size-in	size-out	kernel	param	FLPS
conv1	220×220×3	110×110×64	7×7×3, 2	9K	115M
pool1	110×110×64	55×55×64	3×3×64, 2	0	
rnorm1	55×55×64	55×55×64		0	
conv2a	55×55×64	55×55×64	1×1×64, 1	4K	13M
conv2	55×55×64	55×55×192	3×3×64, 1	111K	335M
rnorm2	55×55×192	55×55×192		0	
pool2	55×55×192	28×28×192	3×3×192, 2	0	
conv3a	28×28×192	28×28×192	1×1×192, 1	37K	29M
conv3	28×28×192	28×28×384	3×3×192, 1	664K	521M
pool3	28×28×384	14×14×384	3×3×384, 2	0	
conv4a	14×14×384	14×14×384	1×1×384, 1	148K	29M
conv4	14×14×384	14×14×256	3×3×384, 1	885K	173M
conv5a	14×14×256	14×14×256	1×1×256, 1	66K	13M
conv5	14×14×256	14×14×256	3×3×256, 1	590K	116M
conv6a	14×14×256	14×14×256	1×1×256, 1	66K	13M
conv6	14×14×256	14×14×256	3×3×256, 1	590K	116M
pool4	14×14×256	7×7×256	3×3×256, 2	0	
concat	7×7×256	7×7×256		0	
fc1	7×7×256	1×32×128	maxout p=2	103M	103M
fc2	1×32×128	1×32×128	maxout p=2	34M	34M
fc7128	1×32×128	1×1×128		524K	0.5M
L2	1×1×128	1×1×128		0	
total				140M	1.6B

TABLE II. THE NN2 INCEPTIONS ARCHITECTURE

type	output size	depth	#1×1	#3×3 reduce	#3×3	#5×5 reduce	#5×5	pool proj (p)	params	FLOPS
conv1 (7×7×3, 2)	112×112×64	1							9K	119M
max pool + norm	56×56×64	0						m 3×3, 2		
inception (2)	56×56×192	2		64	192				115K	360M
norm + max pool	28×28×192	0						m 3×3, 2		
inception (3a)	28×28×256	2	64	96	128	16	32	m, 32p	164K	128M
inception (3b)	28×28×320	2	64	96	128	32	64	L_2, 64p	228K	179M
inception (3c)	14×14×640	2	0	128	256,2	32	64,2	m 3×3,2	398K	108M
inception (4a)	14×14×640	2	256	96	192	32	64	L_2, 128p	545K	107M
inception (4b)	14×14×640	2	224	112	224	32	64	L_2, 128p	595K	117M
inception (4c)	14×14×640	2	192	128	256	32	64	L_2, 128p	654K	128M
inception (4d)	14×14×640	2	160	144	288	32	64	L_2, 128p	722K	142M
inception (4e)	7×7×1024	2	0	160	256,2	64	128,2	m 3×3,2	717K	56M
inception (5a)	7×7×1024	2	384	192	384	48	128	L_2, 128p	1.6M	78M
inception (5b)	7×7×1024	2	384	192	384	48	128	m, 128p	1.6M	78M
avg pool	1×1×1024	0								
fully conn	1×1×128	1							131K	0.1M
L2 normalization	1×1×128	0								
total									7.5M	1.6B

In general, FaceNet is used for training face classifiers with triplet loss function but in this work, we use FaceNet as a face embedding. Each facial image after processed and converted to gray scale should be fed into FaceNet CNNs to get a 128-dimensions embedding vector which represents all the features extracted from the face. All the facial image embedding vectors are saved and used for training the classifier.

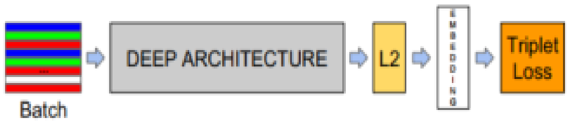


Figure 7. FaceNet model structure [16]

5) Classifier with Support Vector Machine

Although, the FaceNet model is the state-of-the -art in face recognition, verification and clustering neural network with extreme high accuracy but the data set used in this model too large and almost facial images using are European and American so it needs to much computational cost for training and very low accuracy when we apply our data set into it. Alternatively, we apply Support Vector Machine (SVM) [29] algorithms for classifying facial images in this work.

SVM is one of the most popular machine learning algorithms for classification and it is not as complicated as deep learning [30]. Support vector machine aims to find the hyperplane that maximizes the distance between the support vector (the closest data points) and the hyperplane or hypothesis [31]. In other words, there is labeled training the algorithm produces an optimal hyperplane that categorizes new examples. In two dimensional spaces, this hyperplane is a line separating an airplane into two parts where in each class are located on both sides. The hyperplane is visualized in figure 8.

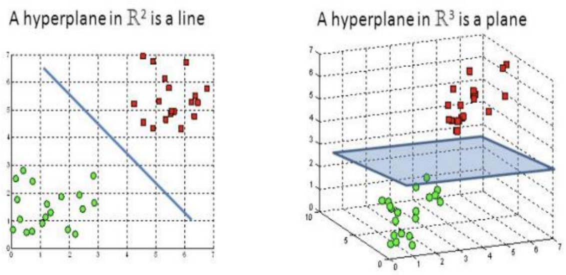


Figure 8. SVM hyperplane

Originally, the SVM algorithm is used to solve some problems classified with only 2 classes, but it can not be solved for multi-classification. So, to implement SVM to multi-classification one of the basic methods called One-Versus-All (1-v-A) has been proposed. In 1-v-A SVM, classification functions are constructed by classifying each class with all other classes [32]. Before doing this, the label of the sample class needs to be replaced by 1 and the labels of other classes replaced by -1. Suppose that you have *n* classes in your dataset. Then, there will be *n* classification needed to be constructed. In the process of classification, the classifier compares all classes, then which class has the largest result will be selected as a recognition class. The loss function of SVM is shown in formula below:

$$f_{loss-SVM} = C\sum_{i=1}^{m}\left[y^{(i)}.Cost_1\left(\theta^T x^{(i)}\right) + \left(1-y^{(i)}\right)Cost_0\left(\theta^T x^{(i)}\right)\right] + \frac{1}{2}\sum_{j=1}^{n}\theta_j^2 \quad (5)$$

where C is penalty parameter, $x^{(i)}$ is sample feature vector, m is number of samples, n is number of features, θ is weight of hypothesis.

6) Text to speech

The smart device after recognizing the face of people in the training dataset should be exported to audio by using a module in python3 named *pyttsx3* for text-to-speech conversion. This module can work offline and have multiple choice for language and accent based on region in many countries in the world.

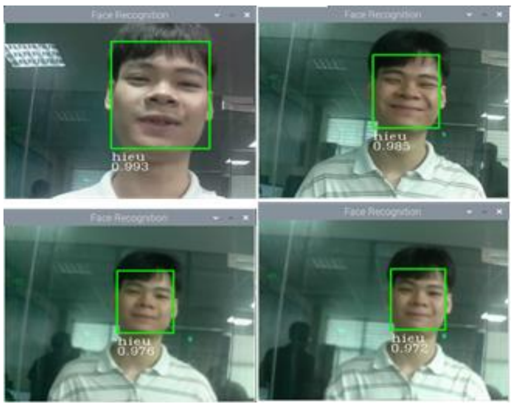
III. SIMULATION RESULTS AND DISCUSSION

After training our model and successfully building our smart wearable device, we surveyed and evaluated the performance of the device in a normal environment with different light conditions. We use accuracy (AC) parameters to measure the person who is identified correctly. Figure 9 shows the prototype of the smart device.

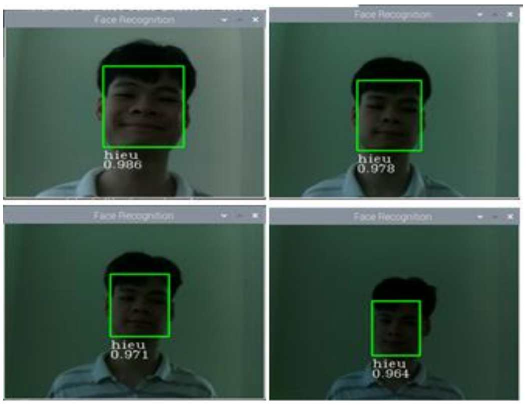
Figure 9. Prototype of the proposed smart device

A. Performance evaluation

The developed smart wearable device needs to ensure accurate face recognition and convey information about the person who is recognized via audio to the visually impaired. With the application of deep learning solutions and optimized classification algorithms, our device ensures high accuracy. Figure 10 shows that the classifier can recognize the human face with high accuracy up to 99.3% in good light and 98.6% even though the device works in weak light condition.



(a)



(b)

Figure 10. Face recognition feature of device in different light condition (a) In good light condition (b) In weak light condition

To investigate the actual capability of the device in different conditions, the research team has experienced the device in different light conditions including: 1) nature light condition (without sunshine and clouds), 2) in a room with weak light condition. Figure 11 describes the accuracy when the smart device recognizes face in two environment conditions above. The results show that the smart device can work well and identify with high accuracy within 140cm. Out of 140cm, the accuracy begins to gradually decrease. The reason for this decrease is currently the research team is using an embedded computer Raspberry Pi 4 with many hardware limitations and processing ability. Besides, the smart device uses a normal Raspberry pi camera with a low resolution of only 5 megapixels. Therefore, we can completely increase the distance identification by using a camera with better sensitivity. However, it is also important to be careful with the balance between performance and cost.

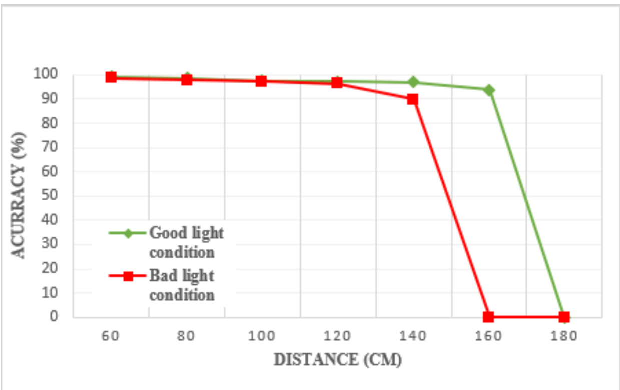


Fig 11. Accuracy of smart device for face recognition in different light condition

B. Discussion

In this work we collected data from only 5 individuals and each person only 300 portrait images with a few poses with the same distance and light condition so it possibly reduces the final recognition performance of the smart device. In order to increase the performance of the device, in the next step, we will collect more data in different environments, distance, light conditions, and research more about pre-processing images before feeding it to the deep learning model.

The computation performance to extract features from people's faces which are detected from the camera also is a huge challenge. In this work we used a raspberry pi with many hardware limitations and processing ability so the smart device still worked a little bit slow. Furthermore, our camera used in this device had low resolution and working distance was not very good so it was hard to capture and detect the face of people from a long distance.

IV. CONCLUSSIONS

The development of smart devices to support visually impaired persons in particular and the disabled in general is considered an urgent issue. Smart devices will help visually impaired people better integrate into life, and be more autonomous with daily activities.

In this paper, we proposed and developed a smart wearable that supports face recognition using deep learning and machine learning classification algorithms, which can be applied in a lot of smart devices. This proposed device archives the great accuracy for face detection and face recognition when using FaceNet as a feature extractor and SVM algorithms used for classification. The proposed smart device aims to create a low-cost device with multiple functions and high performance. It can be widely distributed to individuals and health organizations.

In the near future, we will continue to develop and improve the smart device's functionality to make it more functional and more efficient. Besides, as we mentioned above, with being able to recognize faces using thermal images, now the research team is developing a device with face recognition using thermal images combined with body temperature measurement. We are also currently researching and developing the cash recognition function and will apply into our smart device soon.

REFERENCES

[1] World Health Organization, Visual impairment and blindness: Fact sheet number 282, Aug. 2014. [Online]. Available:http://www.who.int/mediacentre/factsheets/fs282/en/,accessed on: Feb 18, 2022

[2] Z. -Q. Zhao, P. Zheng, S. -T. Xu and X. Wu, "Object Detection With Deep Learning: A Review," in IEEE Transactions on Neural Networks and Learning Systems, vol. 30, no. 11, pp. 3212-3232, Nov. 2019, doi: 10.1109/TNNLS.2018.2876865.

[3] W. Liu, Y. Wen, Z. Yu, M. Li, B. Raj and L. Song, "SphereFace: Deep Hypersphere Embedding for Face Recognition," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 6738-6746, doi: 10.1109/CVPR.2017.713.

[4] L. -H. Lee and P. Hui, "Interaction Methods for Smart Glasses: A Survey," in IEEE Access, vol. 6, pp. 28712-28732, 2018, doi: 10.1109/ACCESS.2018.2831081.

[5] N. M. Kumar, N. Kumar Singh and V. K. Peddiny, "Wearable Smart Glass: Features, Applications, Current Progress and Challenges," 2018 Second International Conference on Green Computing and Internet of Things (ICGCIoT), 2018, pp. 577-582, doi: 10.1109/ICGCIoT.2018.8753047.

[6] N. N. Mohammed, M. I. Khaleel, M. Latif and Z. Khalid, "Face Recognition Based on PCA with Weighted and Normalized Mahalanobis distance," 2018 International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS), 2018, pp. 267-267, doi: 10.1109/ICIIBMS.2018.8549971.

[7] M. Hasan, S. N. H. S. Abdullah and Z. A. Othman, "Face recognition based on opposition particle swarm optimization and support vector machine," 2013 IEEE International Conference on Signal and Image Processing Applications, 2013, pp. 417-424, doi: 10.1109/ICSIPA.2013.6708043.

[8] F. Mahmud, M. T. Khatun, S. T. Zuhori, S. Afroge, M. Aktar and B. Pal, "Face recognition using Principle Component Analysis and Linear Discriminant Analysis," 2015 International Conference on Electrical Engineering and Information Communication Technology (ICEEICT), 2015, pp. 1-4, doi: 10.1109/ICEEICT.2015.7307518.

[9] Zong, Weiwei & Huang, Guang-Bin. (2011). Face recognition based on extreme learning machine. Neurocomputing. 74. 2541-2551. 10.1016/j.neucom.2010.12.041.

[10] X. Ning, W. Li, B. Tang and H. He, "BULDP: Biomimetic Uncorrelated Locality Discriminant Projection for Feature Extraction in Face Recognition," in IEEE Transactions on Image Processing, vol. 27, no. 5, pp. 2575-2586, May 2018, doi: 10.1109/TIP.2018.2806229.

[11] Salloum, Said & Alshurideh, Muhammad & Elnagar, Ashraf & Shaalan, Khaled. (2020). Machine Learning and Deep Learning Techniques for Cybersecurity: A Review. 10.1007/978-3-030-44289-7_5.

[12] P. P. Shinde and S. Shah, "A Review of Machine Learning and Deep Learning Applications," 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), 2018, pp. 1-6, doi: 10.1109/ICCUBEA.2018.8697857.

[13] J. Latif, C. Xiao, A. Imran and S. Tu, "Medical Imaging using Machine Learning and Deep Learning Algorithms: A Review," 2019 2nd International Conference on Computing, Mathematics and Engineering Technologies (iCoMET), 2019, pp. 1-5, doi: 10.1109/ICOMET.2019.8673502.

[14] S. Roy et al., "Deep Learning for Classification and Localization of COVID-19 Markers in Point-of-Care Lung Ultrasound," in IEEE Transactions on Medical Imaging, vol. 39, no. 8, pp. 2676-2687, Aug. 2020, doi: 10.1109/TMI.2020.2994459.

[15] R. S. Latha, G. R. R. Sreekanth, R. C. Suganthe and R. E. Selvaraj, "A survey on the applications of Deep Neural Networks," 2021 International Conference on Computer Communication and Informatics (ICCCI), 2021, pp. 1-3, doi: 10.1109/ICCCI50826.2021.9457016.

[16] F. Schroff, D. Kalenichenko and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 815-823, doi: 10.1109/CVPR.2015.7298682.

[17] Y. Taigman, M. Yang, M. Ranzato and L. Wolf, "DeepFace: Closing the Gap to Human-Level Performance in Face Verification," 2014 IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 1701-1708, doi: 10.1109/CVPR.2014.220.

[18] Zhan Wu, Min Peng and Tong Chen, "Thermal face recognition using convolutional neural network," 2016 International Conference on Optoelectronics and Image Processing (ICOIP), 2016, pp. 6-9, doi: 10.1109/OPTIP.2016.7528489.

[19] Z. -H. Wang, G. -J. Horng, T. -H. Hsu, C. -C. Chen and G. -J. Jong, "A Novel Facial Thermal Feature Extraction Method for Non-Contact Healthcare System," in IEEE Access, vol. 8, pp. 86545-86553, 2020, doi: 10.1109/ACCESS.2020.2992908.

[20] M. Krišto and M. Ivasic-Kos, "An overview of thermal face recognition methods," 2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), 2018, pp. 1098-1103, doi: 10.23919/MIPRO.2018.8400200.

[21] https://www.raspbian.org/

[22] John C. Shovic, "Raspberry pi IoT projects", Berkeley: Apress, 2016.

[23] "Raspberry Pi Foundation". Available online at: https://www.raspberrypi.org/ (Accessed on: Feb. 9, 2022).

[24] N. Agrawal and S. Singhal, "Smart drip irrigation system using raspberry pi and arduino," International Conference on Computing, Communication & Automation, 2015, pp. 928-932, doi: 10.1109/CCAA.2015.7148526.

[25] K. Zhang, Z. Zhang, Z. Li and Y. Qiao, "Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks," in IEEE Signal Processing Letters, vol. 23, no. 10, pp. 1499-1503, Oct. 2016, doi: 10.1109/LSP.2016.2603342.

[26] Z. Yang, W. Ge and Z. Zhang, "Face Recognition Based on MTCNN and Integrated Application of FaceNet and LBP Method," 2020 2nd International Conference on Artificial Intelligence and Advanced Manufacture (AIAM), 2020, pp. 95-98, doi: 10.1109/AIAM50918.2020.00024.

[27] Zeiler M.D., Fergus R. (2014) Visualizing and Understanding Convolutional Networks. In: Fleet D., Pajdla T., Schiele B., Tuytelaars T. (eds) Computer Vision – ECCV 2014. ECCV 2014. Lecture Notes in Computer Science, vol 8689. Springer, Cham. https://doi.org/10.1007/978-3-319-10590-1_53

[28] [28] C. Szegedy et al., "Going deeper with convolutions," 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 1-9, doi: 10.1109/CVPR.2015.7298594

[29] Evgeniou, Theodoros & Pontil, Massimiliano. (2001). Support Vector Machines: Theory and Applications. 2049. 249-257. 10.1007/3-540-44673-7_12.

[30] Marfu'ah, Nur Jati Lantang and Arrie Kurniawardhani. "Comparison of CNN and SVM for Ship Detection in Satellite Imagery." (2020).

[31] M. Akinkunmi, Introduction to Statistics Using R, vol. 11, no. 4. 2019

[32] Zhao Lihong, Song Ying, Zhu Yushi, Zhang Cheng and Zheng Yi, "Face recognition based on multi-class SVM," 2009 Chinese Control and Decision Conference, 2009, pp. 5871-5873, doi: 10.1109/CCDC.2009.5195250.

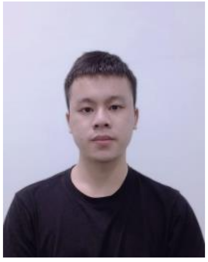
THIẾT BỊ ĐEO THÔNG MINH DÀNH CHO NGƯỜI KHIẾM THỊ DỰA TRÊN THUẬT TOÁN THÔNG MINH

Tóm tắt: Hệ thống thị giác của con người đóng một vai trò quan trọng trong việc nhận thức những thông tin liên quan tới môi trường xung quanh. Bởi vì các tín hiệu thị giác cung cấp nhiều dữ liệu hơn thông tin thính giác nên việc sử dụng những dấu hiệu thị giác hiệu quả hơn sử dụng nhưng tín hiệu thính giác để nhận thức thông tin. Tuy nhiên, trong trường hợp người khiếm thị thiếu thông tin trực quan sẽ hạn chế họ trong việc xác định thông tin và người thân của họ. Do đó, cùng với những công nghệ điều trị từ y tế, công nghệ hỗ trợ người khiếm thị khi di chuyển đang được nghiên cứu và trở nên cấp thiết. Nhận thấy khả năng ứng dụng cao của công nghệ nhận diện với trí tuệ nhân tạo đối với việc hỗ trợ người khiếm thị. Trong bài báo này chúng tôi nghiên cứu và phát triển một thiết bị đeo thông minh nhận dạng khuôn mặt cho người khiếm thị sử dụng kĩ thuật học sâu. Thiết bị được triển khai trên một nền tảng phần cứng gọn nhẹ và linh hoạt với chi phí thấp để nhận diện và thông báo tên người được phát hiện cho người đeo qua giọng nói. Kết quả thử nghiệm được khảo sát và đánh giá qua hiệu năng của thiết bị cho thấy thiết bị vô cùng linh hoạt và đạt độ chính xác xấp xỉ 99.3% trong môi trường thực tế.

Từ khoá: Học sâu, phát hiện khuôn mặt, nhận diện khuôn mặt, mạng nơ-ron tích chập.

Duc Long Vu is currently a B.E. student in Electronics and Telecommunications Engineering of Posts and Telecommunications Institute of Technology (PTIT) of Vietnam. His research interests include cloud computing and machine learning.

Duc Hieu Nguyen is a B.E. student in Electronics and Telecommunications Engineering of Posts and Telecommunications Institute of Technology (PTIT) of Vietnam. His research interests include cloud computing and machine learning.



D.N. Phuong Phi is a B.E. student in Electronics and Telecommunications Engineering of Posts and Telecommunications Institute of Technology (PTIT) of Vietnam. His research interests include IoT and network technologies.

Hai-Chau Le received the B.E. degree in Electronics and Telecommunications Engineering from Posts and Telecommunications Institute of Technology (PTIT) of Vietnam in 2003, and the M.Eng. and D.Eng. degrees in Electrical Engineering and Computer Science from Nagoya University of Japan in 2009 and 2012, respectively. From 2012 to 2015, he was a researcher in Nagoya University of Japan and in University of California, Davis, USA. He is currently a lecturer in Telecommunications Faculty at PTIT. His research interests include optical technologies, network design and optimization and future network technologies. He is an IEEE member.